JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 11, 440-446 (1965)

Remarks on a Generalization of Banach's Principle of Contraction Mappings

SHERWOOD C. CHU*

U.S. Naval Ordnance Laboratory, White Oak, Maryland

AND

J. B. DIAZ†

Institute for Fluid Dynamics and Applied Mathematics, University of Maryland,

and

U.S. Naval Ordnance Laboratory, White Oak, Maryland

Submitted by R. J. Duffin

1. INTRODUCTION

A generalization of Banach's [1; pp. 160-161; Theorem 6] principle of contraction mappings appears in the book of Kolmogorov and Fomin [2]. The purpose of this note is to simplify the proof of the generalization in the book; and to obtain, by proceeding along the lines of the simpler argument, several improvements to the above mentioned generalization (thereby showing that this generalization is merely a special case of an elementary fact).

Section 2 contains a succinct summary of both Banach's principle and its generalization. In Section 3, a simpler proof and several improvements of the generalization of Banach's principle are given. Finally, Section 4 consists of several examples, showing that the results of Section 3 hold for a wider class of transformations than Theorem 2.

2. BANACH'S PRINCIPLE OF CONTRACTION MAPPINGS AND ITS GENERALIZATION

The following exposition of Banach's [1] principle of contraction mappings appears in the book of Kolmogorov and Fomin [2, p. 43]:

* Present address: Bellcomm, Inc., Washington, D.C.

† The research of this author was supported by the Air Force Office of Scientific Research — Grant AFOSR 40063 and by the U.S. Naval Ordnance Laboratory, White Oak, Maryland.

"Let R be an arbitrary metric space. A mapping A of the space R into itself is said to be a *contraction* if there exists a number $\alpha < 1$ such that

$$\rho(Ax, Ay) \leqslant \alpha\rho(x, y), \tag{1}$$

for any two points $x, y \in R$. Every contraction mapping is continuous. In fact if $x_n \to x$, then, by virtue of (1), we also have $Ax_n \to Ax$.

THEOREM 1 (Principle of Contraction Mappings). *Every contraction mapping defined in a complete metric space R has one and only one fixed point (i.e., the equation $Ax = x$ has one and only one solution).*

PROOF. Let x_0 be an arbitrary point. Set $x_1 = Ax_0$, $x_2 = Ax_1 = A^2x_0$, and in general let $x_n = Ax_{n-1} = A^nx_0$. We shall show that the sequence $\{x_n\}$ is fundamental [i.e., Cauchy]. In fact

$$\begin{aligned}\rho(x_n, x_m) &= \rho(A^nx_0, A^mx_0) \leqslant \alpha^n\rho(x_0, x_{m-n})\\ &\leqslant \alpha^n\{\rho(x_0, x_1) + \rho(x_1, x_2) + \cdots + \rho(x_{m-n-1}, x_{m-n})\}\\ &\leqslant \alpha^n\rho(x_0, x_1)\{1 + \alpha + \alpha^2 + \cdots + \alpha^{m-n-1}\} \leqslant \alpha^n\rho(x_0, x_1)(1-\alpha)^{-1}.\end{aligned}$$

Since $\alpha < 1$, this quantity is arbitrarily small for sufficiently large n. Since R is complete, $\lim_{n\to\infty} x_n$ exists. We set $x = \lim_{n\to\infty} x_n$. Then by virtue of the continuity of the mapping A,

$$Ax = A\lim_{n\to\infty} x_n = \lim_{n\to\infty} Ax_n = \lim_{n\to\infty} x_{n+1} = x.$$

Thus, the existence of a fixed point is proved. We shall now prove its uniqueness. If $Ax = x$, $Ay = y$, then $\rho(x, y) \leqslant \alpha\rho(x, y)$, where $\alpha < 1$; this implies that $\rho(x, y) = 0$, i.e., $x = y$."

In the same book of Kolmogorov and Fomin [2, p. 50], there also appears the following generalization of Banach's theorem:

"We note first of all that the principle of contraction mappings can be generalized in the following manner:

THEOREM 2. *If A is a continuous mapping of a complete metric space R into itself, such that the mapping A^n is a contraction for some [positive integer] n, then the equation*

$$Ax = x$$

has one and only one solution.

In fact, if we take an arbitrary point $x \in R$ and consider the sequence $A^{kn}x$ ($k = 0, 1, 2, \cdots$), a repetition of the argument introduced in Section 14 [see Theorem 1 quoted above] yields the convergence of this sequence.

Let

$$x_0 = \lim_{k\to\infty} A^{kn}x.$$

Then

$$Ax_0 \,[= A\,(\lim_{k\to\infty} A^{kn}x)] = \lim_{k\to\infty} A^{kn}Ax.$$

Since the mapping A^n is a contraction, [there is a constant α, with $0 < \alpha < 1$, such that $\rho(A^n x, A^n y) \leqslant \alpha\rho(x, y)$ and] we have

$$\rho(A^{kn}Ax, A^{kn}x) \leqslant \alpha\rho(A^{(k-1)n}Ax, A^{(k-1)n}x) \leqslant \cdots \leqslant \alpha^k\rho(Ax, x).$$

Consequently,

$$\lim_{k\to\infty} \rho(A^{kn}Ax, A^{kn}x) = 0$$

i.e., $Ax_0 = x_0$."

[Clearly, A has a unique fixed point, for if $Ay = y$, then $A^n y = y$, hence $y = x_0$, since A^n has only one fixed point.]

In the above quotations, the numbering of the theorems as 1 and 2, and also the statements within square brackets, are our own.

3. Improvements of the Generalization of Banach's Principle of Contraction Mappings

Remark 1. The proof of Theorem 2 may be simplified somewhat, as follows: Since A^n is a contraction, it possesses, by Theorem 1, a unique fixed point, call it x_0, such that $A^n x_0 = x_0$. It will now be shown that $Ax_0 = x_0$. Since

$$\rho(Ax_0, x_0) = \rho(AA^n x_0, A^n x_0) = \rho(A^n Ax_0, A^n x_0) \leqslant \alpha\rho(Ax_0, x_0),$$

and $\alpha < 1$, one has $\rho(Ax_0, x_0) = 0$, i.e., $Ax_0 = x_0$.

Thus, the argument just given shows that the assumption that A itself is continuous, made among the hypotheses of Theorem 2, is superfluous. However, it should be noticed that the proof of Theorem 2 reproduced in Section 2, nevertheless, does make use of the continuity of A, specifically when it is asserted that, since $x_0 = \lim_{k\to\infty} A^{kn}x$, one has

$$Ax_0 = A(\lim_{k\to\infty} A^{kn}x) = \lim_{k\to\infty} A^{kn}Ax.$$

Therefore, the following theorem is an extension of Theorem 2:

THEOREM 3. *If A is a (single valued) function defined on a complete metric space R into itself, such that the function A^n is a contraction for some (positive integer) n, then A has a unique fixed point.*

REMARK 2. The conclusion that A has a fixed point can be reached in an even more direct manner, still without assuming that A itself is continuous. Since A^n is contracting, it follows from Theorem 1 that A^n has a *unique* fixed point x_0, such that $A^n x_0 = x_0$. Hence

$$Ax_0 = AA^n x_0 = A^n A x_0,$$

which means that Ax_0 is also a fixed point of A^n. But A^n has only one fixed point, and therefore $Ax_0 = x_0$. Thus, x_0 is a fixed polint of A. It is clear that x_0 is a unique fixed point of A, as was shown earlier.

REMARK 3. An examination of the preceding argument shows that there is no need to assume that A^n is contracting and defined on a complete metric space. All that is used in obtaining the conclusion of Theorem 3 is that A^n has exactly one fixed point. Hence one has

THEOREM 4. *Let S be any nonempty set of elements (called "points") and A be a single valued function defined on S and with values in S. Suppose that, for some positive integer n, the function A^n has a unique fixed point x_0. Then A also has a unique fixed point, namely x_0.*

(When S is a complete metric space R, and A is a single valued function on R to R, such that A^n, for some positive integer n, is contracting, then Theorem 4 reduces to Theorem 3.)

REMARK 4. An inspection of the argument leading to the last theorem reveals that the essential property (besides uniqueness of the fixed point for A^n) employed is that A^n and A commute with each other. This suggests immediately the following:

THEOREM 5. *Let S be any nonempty set of elements, and B be a single valued function defined on S and with values in S. Suppose further that B possesses a unique fixed point x_0. Then, if A is* any *single valued function on S to S which commutes with B, that is, such that $AB = BA$, then A also has x_0 as a fixed point (not necessarily unique; however, if B happens to be an iterate of A, that is $B = A^n$, with n a positive integer, then it is unique).*

It should be noticed that Theorem 5 includes Theorems 2, 3, and 4 as special cases.

The proof is immediate, starting from the equation $Bx_0 = x_0$, upon noticing that

$$Ax_0 = ABx_0 = BAx_0.$$

This means that Ax_0 is also a fixed point of B, but B has only x_0 as a fixed point, by hypothesis.

Notice that all that is really used in the above argument is that both x_0 and Bx_0 are in the domain of definition of A, which need not be all of S, and that A and B commute *at* x_0, i.e., that $ABx_0 = BAx_0$.

It should also be noticed that it is precisely the above argument, with $B = A^n$, which is employed in the proof of Theorem 4.

3. Examples

Example 1. This example shows that Theorem 3 is indeed more general than Theorem 2, by displaying a transformation which is not continuous, but whose second iterate is contracting. We are indebted to I. I. Glick for this example. The metric space R is taken to be the Banach space of all real valued continuous functions, $C([0, 1])$, on the closed interval $0 \leqslant x \leqslant 1$, with the norm of a function $f(t)$ being the maximum of $|f(x)|$ for x in this interval. Consider the linearly independent elements (i.e., such that any finite subset is linearly independent) of $C([0, 1])$:

$$e^x, 1, x, x^2, x^3, \cdots,$$

and extend this linearly independent set to a Hamel basis H (i.e., a maximal linearly independent set; see Dunford and Schwartz [3, p. 36]). The transformation A is defined, for elements of H, as follows:

$$A(e^x) = \tfrac{1}{2} \cdot 1, \qquad \text{and} \qquad A(1) = \tfrac{1}{2} \cdot e^x,$$

while $A(h) = \frac{1}{2} h$ for any element of H which is different from 1 or e^x (notice that, therefore, $A(x^n) = \frac{1}{2} x^n$ for $n = 1, 2, \cdots$). Since H is a basis for $C([0, 1])$, the definition of A may be extended, from H to all of $C([0, 1])$, merely by defining $A(y) = \Sigma_{i=1}^n \alpha_i A(h_i)$ whenever $y = \Sigma_{i=1}^n \alpha_i h_i$ (with n a positive integer, real numbers $\alpha_i \neq 0$ for $i = 1, \cdots, n$, and h_i in H for $i = 1, \cdots, n$); further, let $A(0) = 0$. Then $A^2 = \frac{1}{4} I$, where I is the identity mapping. Thus, A^2 is contracting. But A is not continuous at e^x, that is

$$\lim_{n\to\infty} A\left(\sum_{k=0}^{n} \frac{1}{k!} x^k\right) \neq A(e^x) = \tfrac{1}{2},$$

because

$$\begin{aligned}\lim_{n\to\infty} \left[A\left(1 + \sum_{k=1}^{n} \frac{1}{k!} x^k\right)\right] &= \lim_{n\to\infty} \left[\tfrac{1}{2} e^x + \tfrac{1}{2} \sum_{k=1}^{n} \frac{1}{k!} x^k\right] \\ &= \tfrac{1}{2} e^x + \tfrac{1}{2} (e^x - 1) \\ &= e^x - \tfrac{1}{2}.\end{aligned}$$

EXAMPLE 2. It is of interest to notice that an example of a discontinuous transformation A, with A^2 contracting, can be given even when the metric space R is the set of all real numbers. Let the numbers 1 and π be contained in a Hamel basis H for the real numbers (i.e., a set H of rationally independent real numbers such that every nonzero real number may be uniquely written as a finite sum, $y = \Sigma_{i=1}^{n} \alpha_i h_i$, where n is a positive integer, the α_i are nonzero rational numbers, and the h_i are numbers of H; see G. Hamel [4]). The transformation A will be defined, for elements of H, as follows:

$$A(1) = \tfrac{1}{2}\pi, \qquad \text{and} \qquad A(\pi) = \tfrac{1}{2} \cdot 1,$$

while $A(h) = \frac{1}{2} h$ for any number of H which is different from 1 or π. The definition of A may be extended, from H to all the real numbers, by defining $A(y) = \Sigma_{i=1}^{n} \alpha_i A(h_i)$ for any nonzero real number $y = \Sigma_{i=1}^{n} \alpha_i h_i$; and by putting $A(0) = 0$. The transformation A satisfies $A(A(y)) = \frac{1}{4} y$ for every real y, hence A^2 is a contraction. But A cannot be continuous. For, from the way it was defined, A satisfies the Cauchy functional equation

$$A(x) + A(y) = A(x + y).$$

If the function A were continuous, then it would have to be linear, that is

$$A(y) = cy,$$

for some real number c, and any real y. Since $c = A(1) = \pi/2$, one would then have that $A(\pi) = c \cdot \pi = \pi^2/2$, contradicting the original definition of A which states that $A(\pi) = \frac{1}{2}$.

EXAMPLE 3. This example shows the "power" of Theorem 5, as compared with the preceding theorems. This is done by displaying a transformation A which is not contracting, not continuous, and such that no iterate of A is contracting. Nevertheless, it may be shown, "strictly" as a consequence of Theorem 5, that A has a fixed point, while this conclusion cannot be inferred as a consequence of the previous Theorems 1 to 4 (since A has, obviously, as will be seen from its definition, more than one fixed point). The set S is taken to be the set of all finite complex numbers $z = x + iy$, with x and y real. The function A is defined as follows:

$$A(z) = 1/\bar{z};$$

for z not zero, where $\bar{z} = x - iy$ is the complex conjugate of the number z (so far, A is essentially a "Kelvin" inversion); further, let $A(0) = 0$. It is not a surprise that the function A, obviously, has zero, and the set of numbers of absolute value one, as its fixed points. Hence, it is not possible to "deduce," strictly as a consequence of any of Theorems 1-4, that A has a fixed point. However, if B is a rotation of the complex number plane about the origin, through an angle which is not an integral multiple of 2π, then B has zero as its only fixed point, and B commutes with A. Hence, it is strictly possible to "deduce" that zero is also fixed point of A.

Note added in proof. This note appeared first as a United States Naval Ordnance Report NOLTR 64-39, on March 2, 1964, and was accepted for publication on April 7, 1964. After reading the report, Professor I. I. Kolodner kindly communicated to us, in a letter of May 11, 1964, that he had known our Theorem 4 for some time, and that he had submitted the result to the American Mathematical Monthly; it has subsequently appeared as one of the Classroom Notes (*Am. Math. Monthly* **71**, No. 8 (October 1964), 906).

References

1. S. Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales (Thèse présentée en juin 1920 a l'Université de Léopol [= Lwów] pour obtenir le grade de docteur en philosophie). *Fundamenta Math.* **3** (1922), 133-181.
2. A. N. Kolmogorov, and S. V. Fomin. "Elements of the Theory of Functions and Functional Analysis, Volume I, Metric and Normed Spaces." Graylock Press, Rochester, New York, 1957.
3. N. Dunford, and J. T. Schwartz. "Linear Operators, Part I, General Theory." Interscience, New York, 1958.
4. G. Hamel. Eine Basis aller Zahlen und die unstetige Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$. *Math. Ann.* **60** (1905), 459-462.